EDA ile Kategorik Değişken Analizi 1 (Analiysis of Categorical)

Yine Veri Setimiz küçük olduğunda bu şekilde bir anlayışla analiz yapmak kolaydır.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

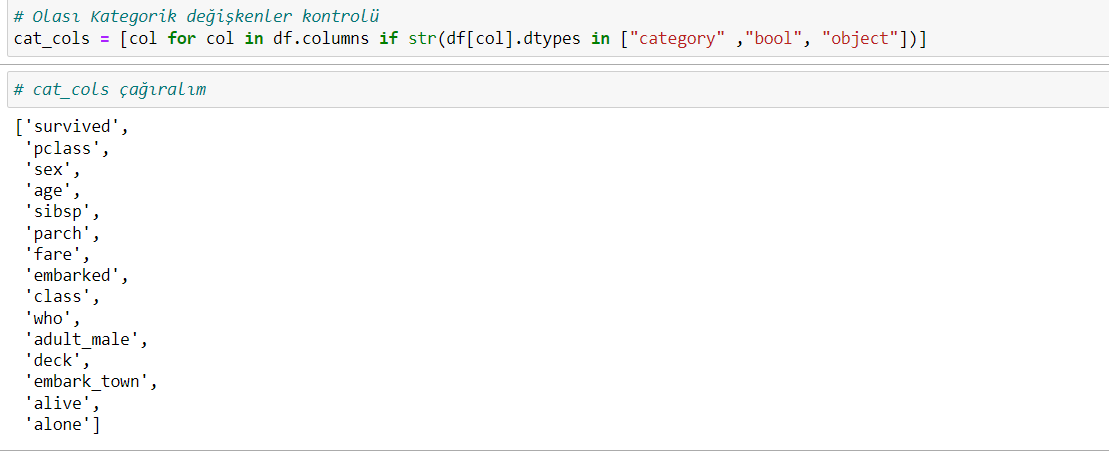
NOT: nunique = “number unique” yani eşsiz değerlerin adedini belirtmek içindir. (Değişken Sınıfı => Değişken Sınıf Sayısı)

Fakat daha büyük milyonluk veri setlerinde bu şekilde bir anlayışla analiz yapmak mümkün değildir.

EDA ile Kategorik Değişken Analizinin amacı, kategorik değişkenleri yakalamaktır.

Yapılması gerekenler:

1. Değişkenleri type’lerine göre analiz etmeliyiz
2. Type’leri Sayısal olsa dahi gizlenmiş kategorik değişkenleri yakalamayız
3. Typeleri kategorik görünen ama aslında kategorik olmayan değişkenleri de yakalamayız
4. Maddenin Örnek Kullanımı



1. Maddenin Örnek Kullanımı

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

1. Maddenin Örnek Kullanımı

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Bu işlemlerin sonunda ise

* Kategorik değişkenleri bir araya toplamlıyız,
* kategorik görünen ama aslında sayısal olanları da çıkarmalıyız ki temiz bir sonucu elde tutabilelim.

1. İlk olarak tüm kategorikleri bir araya toplayalım

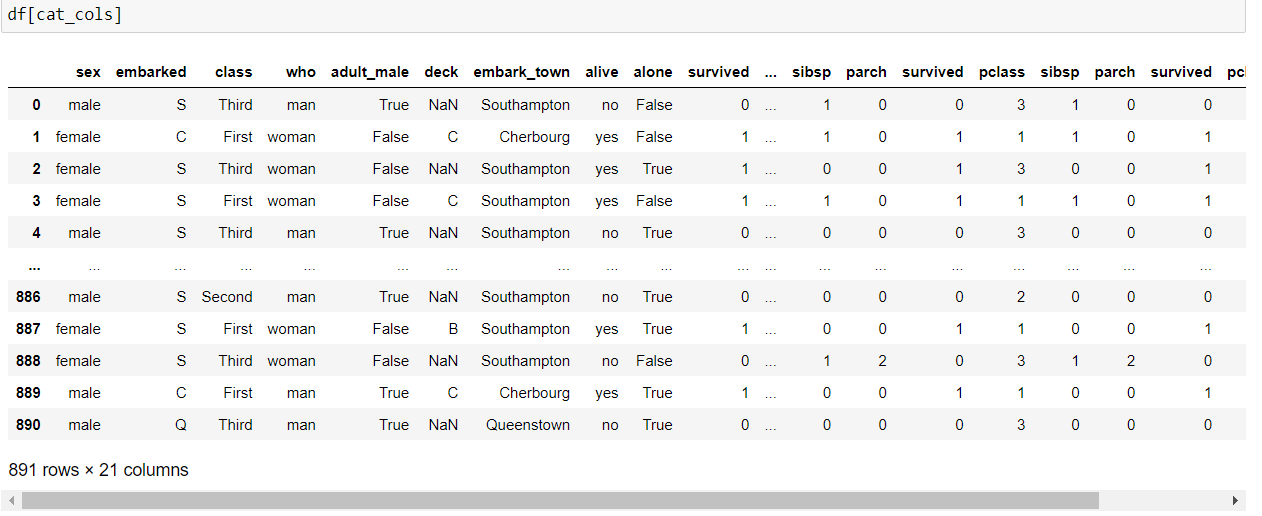


1. Kategorik görünen ama aslında sayısal (cardinal) olanları çıkaralım



Bu işlemler sonucunda tüm kategorik değişkenleri çağıralım

NOT: Bu yaptığımız işlemler Model Deployment ve Development işlemlerinde, büyük ölçekli Veri Bilimi ve Veri Analitiği süreçlerinde kritik öneme sahiptir.



ÖNEMLİ !!!

Tüm yapılan işlemlerin sağlamasını yapmak ve kontrol etmek istersek



metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Elimizdeki değişkenlerin unique (eşsiz) olmayan değerlerinden yola çıkarak bunların kategorik değişken olduğunu söyleyebiliriz.

Peki bir diğer sağlama olarak Sayısal Değişkenlere bakmak istersek? (cat\_cols’un içinde olmayan değerleri seçelim)





Output’lara baktığımızda görüyoruz ki yaptığımız işlemler doğru sonuç veriyor. (Elde edilen değişkenler sayısal (age, fare))

SONUÇ !!!

Yapılan işlemlerin sonucu olarak artık Fonksiyonel Analiz çerçevesinde bir işleme geçebiliriz.

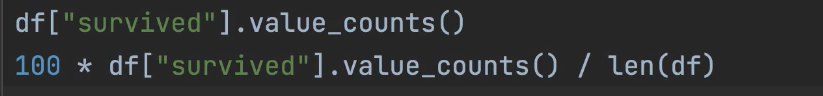
İlk olarak değişkenlerin value\_counts değerlerine bakacağız

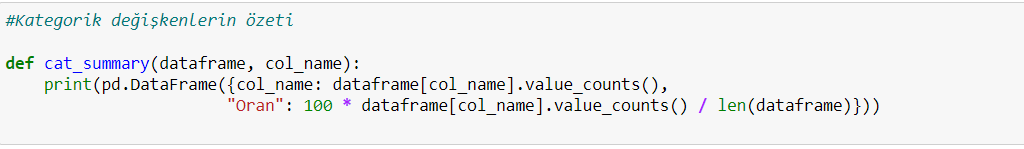
Value count = DEĞER SAYISI

Daha sonra da değişkenlerin oransal olarak bir değerlendirmesine bakacağız (100 \* uzunluk)

Bir fonksiyon oluşturup, bu işlemleri oraya aktaracağız.

Örnek





metin içeren bir resim

Açıklama otomatik olarak oluşturuldu



Artık veri setimizdeki tüm kategorik değişkenler için bir for döngüsü oluşturarak fonksiyonumuzu tüm kategorik değişkenlere uygulayabilir ve oransal (yüzdelik) değerleri görebiliriz.

**NOT**: Şimdilik cat\_cols üzerinden işlemler yaparak kategorik değişkenleri işleme tabii tuttuk

Fakat kategorik değişkenleri seçerken kullandığımız lambda ifadelerini de bir fonksiyon içerisine atabilir ve onları da fonksiyonel programlamaya dahil edebiliriz.

Diğer çalışmalarda ileri seviye fonksiyonel işlemler devam edecektir…